

DEVICE AND SYSTEM FOR UNDERSTANDING VOICE WORD

Patent number: JP11143493
Publication date: 1999-05-28
Inventor: MASATAKI HIROKAZU
Applicant: ATR ONSEI HONYAKU TSUSHIN KENKYUSHO:KK
Classification:
 - **international:** G10L3/00; G10L3/00; G10L3/00; G06F15/18;
 G06F17/28; G06F17/30
 - **european:**
Application number: JP19970303075 19971105
Priority number(s):

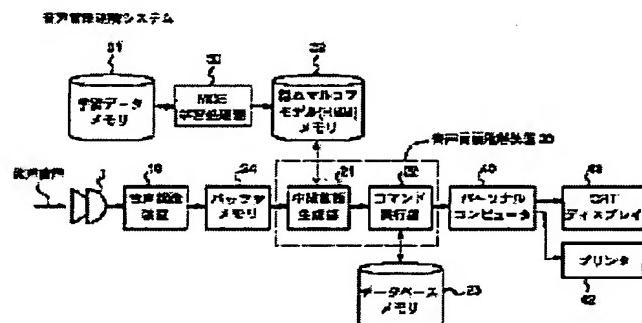
Also published as:

JP11143493 (/)

Abstract of JP11143493

PROBLEM TO BE SOLVED: To provide a device and a system for understanding the voice word capable of firmly and correctly understanding the voice, and executing the appropriately responding processing compared with the conventional technology.

SOLUTION: An MCE learning part 30 learns a hidden Markov model to convert sentence data into intermediate words corresponding thereto so that the discrimination error is minimized based on the learned data. An intermediate word generation part 21 converts the voice sentence of the result of the speech recognition which is speech recognized and includes the retrieval condition using a hidden Markov model to convert the sentence data into the intermediate words corresponding thereto. After a command execution part 22 converts the generated intermediate words into the prescribed database words corresponding to the database, it retrieves the database based on the retrieval condition included in the database words, acquires the data to satisfy the retrieval condition, and executes the processing responding to the intermediate words on the data.

Data supplied from the esp@cenet database - Patent Abstracts of Japan**BEST AVAILABLE COPY**

JARZ-10-PCT

(ISR) 1,

(19) 日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平11-143493

(43)公開日 平成11年(1999)5月28日

(51) Int.Cl. ⁶	識別記号
G 1 O L 3/00	5 7 1
	5 3 5
	5 6 1
G O 6 F 15/18	5 6 0
	17/28

F I		
G 1 O L	3/00	5 7 1 H
		5 3 5
		5 6 1 G
G 0 6 F	15/18	5 6 0 G
	3/16	3 2 0 H

審査請求 有 請求項の数 2 O.L (全 14 頁) 最終頁に統ぐ

(21) 出願番号 特願平9-303075

(22) 出願日 平成 9 年(1997)11 月 5 日

(71)出願人 593118597

株式会社エイ・ティ・アール音声翻訳通信
研究所

研究所
京都府相楽郡精華町大字乾谷小字三平谷5
番地

(72) 春明者 政濶 造和

京都府相楽郡精華町大字乾谷小字三平谷5
番地 株式会社エイ・ティ・アール音声翻
訳通信研究所内

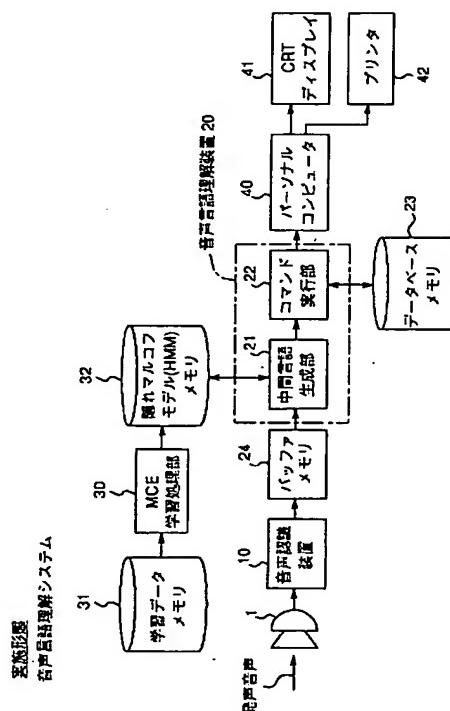
(74)代理人 弁理士 青山 蕊 (外2名)

(54) 【発明の名称】 音声言語理解装置及び音声言語理解システム

(57) 【要約】

【課題】 従来技術に比較して頑健にかつ正確に音声理解を行うことができ、適切に応答する処理を実行することができる音声言語理解装置及び、音声言語理解システムを提供する。

【解決手段】 MCE学習処理部30は学習データに基づいて識別誤りが最小となるように文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習する。中間言語生成部21は、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを用いて、音声認識されかつ検索条件を含む音声認識結果の音声文を中間言語に変換して生成する。コマンド実行部22は、生成された中間言語を、データベースに対応した所定のデータベース言語に変換した後、データベース言語に含まれる検索条件に基づいて、データベースを検索して、検索条件を満たすデータを獲得し、そのデータについて中間言語に対応した応答する処理を実行する。



【特許請求の範囲】

【請求項1】 発声音声から音声認識装置によって音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解装置であつて、

検索条件を含む文章データと、それに対応しかつ少なくとも応答する処理の内容及びデータの項目名を含む所定の中間言語データとの対である学習データに基づいて、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習して得られた隠れマルコフモデルを記憶する第1の記憶装置と、

複数の項目名に対応したデータを含むデータベースを記憶する第2の記憶装置と、

上記第1の記憶装置に記憶された隠れマルコフモデルを用いて、上記音声認識されかつ検索条件を含む音声認識結果の音声文を上記中間言語に変換して生成する生成手段と、

上記生成手段によって生成された中間言語を、上記データベースに対応した所定のデータベース言語に変換した後、上記データベース言語に含まれる検索条件に基づいて、上記第2の記憶装置に記憶されたデータベースを検索して、上記検索条件を満たすデータを獲得し、そのデータについて上記中間言語に対応した応答する処理を実行する実行手段とを備えたことを特徴とする音声言語理解装置。

【請求項2】 発声音声を音声認識して、音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解システムであつて、

発声音声を音声認識して、音声認識された音声認識結果の音声文を出力する音声認識装置と、

検索条件を含む文章データと、それに対応しかつ少なくとも応答する処理の内容及びデータの項目名を含む所定の中間言語データとの対である学習データに基づいて、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習して得られた隠れマルコフモデルを記憶する第1の記憶装置と、

複数の項目名に対応したデータを含むデータベースを記憶する第2の記憶装置と、
上記第1の記憶装置に記憶された隠れマルコフモデルを用いて、上記音声認識装置によって音声認識されかつ検索条件を含む音声認識結果の音声文を上記中間言語に変換して生成する生成手段と、
上記生成手段によって生成された中間言語を、上記データベースに対応した所定のデータベース言語に変換した後、上記データベース言語に含まれる検索条件に基づい

て、上記第2の記憶装置に記憶されたデータベースを検索して、上記検索条件を満たすデータを獲得し、そのデータについて上記中間言語に対応した応答する処理を実行する実行手段とを備えたことを特徴とする音声言語理解システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、発声音声から音声認識装置によって音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解装置、並びに、音声認識装置及び音声言語理解装置とを備えた音声言語理解システムに関する。

【0002】

【従来の技術及び発明が解決しようとする課題】近年、隠れマルコフモデルを用いた音響モデル、及びN-gramを用いた言語モデルを用いた連続音声認識が盛んに研究されており、数万語彙の認識でも、単語認識率が90%以上とかなり実用レベルに近くなっている。しかしながら、音声認識技術を用いたアプリケーションを考えた場合、読み上げた文章をそのまま出力するディクテーションシステムを除くと、旅客機案内システム、電話番号案内システム、音声翻訳システム等、音声認識結果を理解し、ユーザーに情報を提供するいわゆる「音声理解システム」の方が応用分野が広いと考えられる（例えば、従来技術文献1「坂井信輔ほか、『音声入力を用いたパソコンネットワーク旅客機空席案内システムの試作』、電子情報通信学会技術報告、SP94-89、p. 29-36, 1995年1月」参照。）。

【0003】現在、音声理解システムのための言語理解の技術は、発話の内容を構文で限定したものや文法理論を用いたもの（例えば、従来技術文献2「S. Seneff, "TINA: A Natural Language System for Spoken Language Applications", Computational Linguistics, Vol. 18, No. 1, 1992年3月」参照。）が主流である。発話内容を構文で限定する手法は、理解率は高いと考えられるが、計算機が受理できる発話内容をユーザーが事前に知っていなければならず、ユーザーへの負担が大きく、使いやすいシステムとは言えない。

【0004】また、文法理論を用いた方法は、構文による手法よりは、発話内容の自由度が高いが、文法的に正しい文章でないと、理解できないという問題がある。しかしながら、音声認識で広く使われているN-gram言語モデルは、認識率の観点からは非常に有利とされているが、直前の(N-1)単語から次の単語への接続を確率で表現するという極めて単純なモデルであるため、局所的な制約しか表現できず、文全体として必ずしも文

法的に正しい文章を出力するとは限らない。従って、認識結果に誤りが含まれる場合、正しく言語理解を行うのは困難であると考えられる。また、実際のシステムの使用時には、ユーザーが文法的に正しい文章を発声することは限らず、自然発話を理解するのは困難である。

【0005】この問題を解決するため、認識結果文を言語理解部が受理できる部分に分割する手法等（例えば、従来技術文献3「Y. Wakita et al.,

"Correct parts extraction from speech recognition results using semantic distance calculation, and its application to speech translation", ACL, 1997年」参照。）が提案されているが、分割を行うことにより、文章の大局的な情報を得ることができないと考えられる。

【0006】また、従来技術文献4「遠藤充ほか、「音声による文例検索システムの検討」、日本音響学会講演論文集, 2-Q-12, pp. 163-164, 1997年3月」においては、キーワードによる方法が提案されているが、キーワードのみでは文章の意味を正しく理解することができず、また、ユーザーインターフェース等でキーワードの間を補う必要がある。

【0007】本発明の目的は以上の問題点を解決し、上記従来技術に比較して頑健にかつ正確に音声理解を行うことができ、適切に応答する処理を実行することができる音声言語理解装置及び、音声言語理解システムを提供することにある。

【0008】

【課題を解決するための手段】本発明に係る請求項1記載の音声認識装置は、発声音声から音声認識装置によって音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解装置であって、検索条件を含む文章データと、それに対応しかつ少なくとも応答する処理の内容及びデータの項目名を含む所定の中間言語データとの対である学習データに基づいて、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習して得られた隠れマルコフモデルを記憶する第1の記憶装置と、複数の項目名に対応したデータを含むデータベースを記憶する第2の記憶装置と、上記第1の記憶装置に記憶された隠れマルコフモデルを用いて、上記音声認識されかつ検索条件を含む音声認識結果の音声文を上記中間言語に変換して生成する生成手段と、上記生成手段によって生成された中間言語を、上記データベースに対応した所定のデータベース言語に変換した後、上記データベース言語に含まれる検索条件に基づいて、上記第2の記憶装置に記憶されたデータベースを検索して、上記検索条件を満たすデータを獲得し、そのデータについて上記中間言語に対応した応答する処理を実行する実行手段とを備えたことを特徴とする。

そのデータについて上記中間言語に対応した応答する処理を実行する実行手段とを備えたことを特徴とする。

【0009】また、本発明に係る請求項2記載の音声言語理解システムは、発声音声を音声認識して、音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解システムであって、発声音声を音声認識して、音声認識された音声認識結果の音声文を出力する音声認識装置と、検索条件を含む文章データと、それに対応しかつ少なくとも応答する処理の内容及びデータの項目名を含む所定の中間言語データとの対である学習データに基づいて、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習して得られた隠れマルコフモデルを記憶する第1の記憶装置と、複数の項目名に対応したデータを含むデータベースを記憶する第2の記憶装置と、上記第1の記憶装置に記憶された隠れマルコフモデルを用いて、上記音声認識装置によって音声認識されかつ検索条件を含む音声認識結果の音声文を上記中間言語に変換して生成する生成手段と、上記生成手段によって生成された中間言語を、上記データベースに対応した所定のデータベース言語に変換した後、上記データベース言語に含まれる検索条件に基づいて、上記第2の記憶装置に記憶されたデータベースを検索して、上記検索条件を満たすデータを獲得し、そのデータについて上記中間言語に対応した応答する処理を実行する実行手段とを備えたことを特徴とする。

【0010】

【発明の実施の形態】以下、図面を参照して本発明に係る実施形態について説明する。

【0011】図1は、本発明に係る一実施形態である音声言語理解装置20を備えた音声言語理解システムの構成を示すブロック図である。本発明に係る実施形態においては、統計的処理に基づく中間言語への変換を行うことにより、ユーザーの発声に対して頑健な理解が行える音声言語理解装置20を提供することを特徴としている。

【0012】ここで、音声言語理解装置20は、発声音声から音声認識装置によって音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解装置であり、(a) 学習データメモリ31に格納され、検索条件を含む文章データと、それに対応しかつ少なくとも応答する処理の内容及びデータの項目名を含む所定の中間言語データとの対である学習データに基づいて、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデル(HMM)をMCE学習処理部30によって学習して得られた隠れマルコフモデルを記憶する隠れマルコフモデルメモリ32と、(b) 複数の

項目名に対応したデータを含むデータベースを記憶するデータベースメモリ23と、(c)隠れマルコフモデルメモリ32に記憶された隠れマルコフモデルを用いて、上記音声認識されかつ検索条件を含む音声認識結果の音声文を上記中間言語に変換して生成する中間言語生成部21と、(d)中間言語生成部21によって生成された中間言語を、上記データベースに対応した所定のデータベース言語に変換した後、上記データベース言語に含まれる検索条件に基づいて、データベースメモリ23に記憶されたデータベースを検索して、上記検索条件を満たすデータを獲得し、そのデータについて上記中間言語に対応した応答する処理、具体的には表示処理を実行するコマンド実行部22とを備えたことを特徴としている。

【0013】まず、本実施形態の音声言語理解システムの概要について説明する。本実施形態では、音声言語理解システムとして、音声入力による指示により、データベースメモリ23内データベースへアクセスし、ユーザーの要求する情報を表示するシステムを開示する。好ましい実施形態として構築したシステムは、スキー場案内システムであって、音声により、スキー場のデータの入ったデータベースにアクセスし、必要な情報を得るシステムである。システム全体の構成を図1に示す。本システムは、主に「音声認識装置10」と「音声言語理解装

置20」とで構成される。

【0014】音声認識装置10では、入力された発聲音声の波形データに対して特徴量計算を行った後、公知の隠れマルコフ網による音響モデル、及び公知の可変長N-g r a mによる言語モデルを用いて、単語グラフサーチ法により解の探索を行い、認識結果を出力する。音声認識装置10の認識結果は音声言語理解装置20に渡される。音声言語理解装置20では、音声認識結果の単語列を、中間言語に変換し、中間言語の内容に基づいてデータベース言語の生成し、データベースから情報の検索を行い、中間言語の要求に応じて検索結果の表示を行う。本システムは、次の3つの動作を行うことができる。

(a) 各スキー場のデータ（県・標高差・リフト数等12項目）の表示（SHOWVALUE）、(b) ユーザーが要求する条件を満たすスキー場の検索（SHOWLIST）、及び(c) スキー場の地図の表示（SHOWIMAGE）。

【0015】次いで、音声言語理解装置20について説明する。その動作の概要を表1に示す。また、データベースメモリ23内のデータベースの一例を表2に示す。

【0016】

【表1】

入力文：“八方尾根スキー場の標高差を教えて下さい”

1. 中間言語生成

“R_SHOWVALUE O_標高差 D_スキー場名 C_= V_八方尾
根”

2. データベース言語への変換

“SELECT 標高差 FROM スキー場データ
WHERE スキー場名=八方尾根”

【0017】表1の2.においてデータベース言語の一例について示している。この例では、「スキー場データ」というデータベース名のデータベースから(FROM)、「スキー場名=八方尾根」という検索条件のもと(WHERE)で、項目名「標高差」のデータを検索

データベースの一例

スキー場データ

(SELECT) しなさいというデータベース言語である。

【0018】

【表2】

スキー場名	県	標高差	入場者数	
志賀高原	長野	500	1000000	
野沢温泉	長野	1100	900000	
妙高赤倉	新潟	800	800000	
八方尾根	長野	1000	700000	←動作例
梅池高原	長野	700	600000	

(注) 動作例

3. スキー場名=八方尾根の行を検索

4. 標高差を出力

【0019】音声言語理解装置20の処理は、音声認識結果をデータベースアクセス用の中間言語に変換することにより行う。本システムで用いた中間言語は次の要素から構成される。

(a) R_ (コマンド名)

要求動作の指定 (Request)

(b) O_ (対象物名)

動作の対象 (Object)

R_ (コマンド名) O_ (対象物名1) ... O_ (対象物名m)

D_ (ドメイン名1) C_ (比較方法1) V_ (値1)

...

D_ (ドメイン名n) C_ (比較方法n) V_ (値n)

【0022】以下に、自然言語から中間言語への変換例を挙げる。以下の変換例は、図1の学習データメモリ31に、学習データの文章データとそれに対応する中間言語データとして格納される。

(a) 八方尾根スキー場の標高差はいくらですか。

→R_SHOWVALUE O_ 標高差

D_ スキー場名 C_ = V_ 八方尾根

(b) 標高差が1000m以上のスキー場を教えて下さい。

→R_SHOWLIST O_ スキー場名 D_ 標高差
C_ >= V_ 1000

(c) 八方尾根のゲレンデマップを見せて下さい。

→R_SHOWIMAGE O_ ゲレンデマップ

D_ スキー場名 C_ = V_ 八方尾根

【0023】音声言語理解装置20の一連の動作を表1及び表2に示す。音声言語理解装置20は、音声認識結果が入力されると、次の順序で処理を行う。

(1) 音声認識結果から中間言語への変換処理、(2) 中間言語の、対象物名、ドメイン名(表2における項目名である)、比較方法、及び比較値からデータベース言語を生成する処理、(3) 条件に適合するデータをデータベースから検索し、動作の対象情報を獲得する処理、及び、(4) 対象物名に対して中間言語のコマンド名で規定された動作を実行する処理。ここで、上記

(1) の処理は図1の中間言語生成部21によって実行

$$\underset{S}{\operatorname{argmax}} P(S|W) = \underset{S}{\operatorname{argmax}} P(W|S) P(S)$$

上記数2で、確率 $P(W|S)$ は、中間言語から音声認識結果が出力される確率を意味する。この確率を直接的に求めるのは困難なため、次式の近似を考える。

【数3】

(c) D_ (ドメイン名)

データベースの検索項目 (Domain)

(d) C_ (比較方法)

データベース検索時の比較方法 (Comparison)

(e) V_ (値)

データベース検索時の比較値 (Value)

【0020】中間言語は、これらの要素の列として表現され、次の表で示すフォーマットで与えられる。

【0021】

【表3】

され、上記(2)、(3)及び(4)の処理は図1のコマンド実行部22によって実行される。なお、データベース言語は公知のSQL言語と類似した言語を用いており、中間言語は、データベース言語へ必ず正しく変換されるように設計されている。

【0024】次いで、自然言語から中間言語への変換について説明する。音声言語理解装置10において、最も重要かつ困難な部分は、音声認識の入力文章から中間言語への変換部分である。本実施形態は、これを統計的手法に基づいて実行する方法を用いる。

【0025】統計的手法による自然言語から中間言語への変換では、単語系列Wが与えられたとき、次式を満たす中間言語列Sを求めることにより、最適な中間言語を得ることができる。

【数1】

$$\underset{S}{\operatorname{argmax}} P(S|W)$$

ここで、 $P(S|W)$ は、単語系列Wが与えられたときの中間言語列Sを取り得る確率であり、数1は、中間言語列Sを変化したときに確率 $P(S|W)$ が最大となるときの中間言語列Sを表わす。

【0026】上記数1は、ベイズ則を用いると、次式のように表される。

【数2】

$$P(W|S) = \sum_i P(W|s_i)$$

【0027】すなわち、中間言語のそれぞれの要素は独立しており、また、中間言語のそれぞれの要素は、特定の単語のみを出力するのではなく、入力文の全ての単語がある確率で出力すると仮定する。この仮定により、誤

認識が生じた際や、不完全な文が入力された場合でも、中間言語への変換が容易になり、頑健な言語理解が可能になる。この確率 $P(W|S)$ を求めるモデルを文生成モデルと呼ぶ。一方、確率 $P(S)$ は、中間言語の事前確率で、入力文とは全く独立に求められる。統計的手法による自然言語から中間言語への変換の精度は、これらの確率の精度に依存する。

【0028】次いで、これらの確率を求めるための文生成モデルについて述べる。上記確率 $P(W|s_i)$ 、即ち、中間言語の各要素 s_i から音声認識結果文を与えるモデルとして、隠れマルコフモデル (HMM) を用いる（図4参照）。隠れマルコフモデルは、図4に示すように、複数の状態から構成され、単語が入力される毎に、状態 i から状態 j へ確率 a_{ij} で遷移し、遷移後の状態 j から確率 b_{j,w_t} で単語を出力するモデルである。隠れマルコフモデルは、音声認識の音響モデルにおいては、通常 Left-to-Right型のモデルが用いられるが、ここで用いるモデルは、全ての状態間遷移が可能なエルゴディックモデルを考える。ここで、用いる隠れマルコフモデル (HMM) のパラメータは次の通りである。

- (a) 状態数 : M
- (b) 状態 i から状態 j への遷移確率 : a_{ij}
- (c) 状態 j から単語 w_t への出力確率 : b_{j,w_t}

【0029】隠れマルコフモデルのパラメータの推定は、文章とそれに対応する中間言語列のデータを大量に容易し、 $P(W|s_i)$ の値が最大になるように決定する（最尤推定法による）。隠れマルコフモデルの場合、観測データに対応する状態系列が分からぬため、公知のEM (Expectation-Maximum) アルゴリズムによって最尤推定を行う。隠れマルコフモデルの場合は特に、バーム・ウェルチ (Baum-Welch) アルゴリズムと呼ばれる。隠れマルコフモデルは、中間言語の各要素毎に作成し、文が入力されると、全てのモデルが独立に、平行して動作する。隠れマルコフモデルを用いて、入力文から中間言語への変換を行な際は、公知のビタビ (Viterbi) アルゴリズムを用いてそれぞれの要素に対して文の生成確率のみを求め、R_, O_, D_, C_, V_のそれぞれのグループ内で最も確率の高い要素を選び、中間言語列を得る

$$d = -g_k(X, \Lambda) + \log(\{1/(K-1)\})$$

$j, j \neq k$

ここで、 k は読み込んだデータの中間言語に含まれる要素で、 j は k のグループに属する中間言語の要素である。（c）損失関数

【数6】

$$l(d_k) = 1 / (1 + \exp[-a(d_k + b)])$$

ここで、 a 及び b は予め経験的に決定される定数である。

（図5参照。）。

【0030】次いで、中間言語への変換の高精度化のための、図1のMCE学習処理部30によって実行される公知のMCE (Minimum Classification Error; 識別誤り最小法) トレーニングによる学習について説明する。通常、隠れマルコフモデルの学習は、公知のEMアルゴリズムによって行われる。EMアルゴリズムは、基本的には最尤推定法であり、本実施形態で用いる文生成モデルに使用した場合、不都合が生じる場合がある。例として、次の2つの場合について考える。（1）長野県のスキー場を教えて下さい。

→ R_SHOWLIST O_スキー場名 D_県 C = V_長野

（2）長野県以外のスキー場を教えて下さい。

→ R_SHOWLIST O_スキー場名 D_県 C <> V_長野

【0031】この場合、文（1）と文（2）との差は、単語「以外」があるかないかのみの違いであり、その差が中間言語の「C_=」と「C_<>」との差となる。しかしながら、上述の文生成モデルであると、「C_<>」のモデルでは、文（1）例に出現する単語全てに対して比較的高い確率を出力し、（1）の文が入力された場合「C_=」と「C_<>」との区別が困難になる可能性がある。このため、類似した文に対する識別度を向上させるため、隠れマルコフモデルに対して、MCE学習を行う。

【0032】MCEに基づく学習法は、クラスの識別に用いる尺度を識別関数 g_k とし、あるサンプル X に対する識別関数の差で表される識別誤り関数 $d_k(X, \Lambda)$ から、シグモイド (sigmoid) 関数で現れる損失関数 $l(d_k)$ を用いて実効的な識別誤り数を評価し、この識別誤り数を最小化する基準でモデルパラメータ Λ を求める方法である。ここで、用いる識別関数、識別誤り関数、及び損失関数を次式に示す。

【0033】（a）識別関数

$$【数4】 g(X, \Lambda) = \log[L(x)]$$

ここで、 $L(x)$ は、隠れマルコフモデルによる文生成確率 $P(W|s_i)$ を表わす。（b）識別誤り関数

【数5】

$$\Sigma_{j, j \neq k} \exp[\eta g_j(X, \Lambda)]^{1/\eta}$$

【0034】損失関数 $l(d_k)$ に対して、最急降下法を用いて漸化的にパラメータ Λ を調整しながら、最適パラメータを求める。

$$【数7】 \Lambda_{h+1} = \Lambda_h - \epsilon \nabla l(d_k(X; \Lambda_h))$$

【0035】図6は、図1のMCE学習処理部30によって実行されるMCE学習処理を示すフローチャートである。MCE学習処理部30は、学習データメモリ31

内の学習データの文章データとそれに対応する中間言語データに基づいて、以下のMCE学習処理を実行することにより、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習して隠れマルコフモデルメモリ32に格納する。

【0036】図6において、まず、ステップS0でパラメータ i に1を代入し、ステップS1で学習データメモリ31から、文章データ及び中間言語データをそれぞれ1文読み込む。次いで、ステップS2で、数6を用いて、損失 L を計算し、ステップS3で数7を用いて隠れマルコフモデル(HMM)の各パラメータを更新する。そして、ステップS4で処理すべき学習データがまだあ

るか否かが判断され、YESのときはステップS1に戻り上記の処理を繰り返す一方、ステップS5では全ての文について総損失 L を計算する。そして、ステップS6では、学習終了判定が実行されて、 $|L_i - L_{i-1}| < C$ (ここで、Cは定数である。)であるか否かが判断され、NOのときは終了していないとして、ステップS7でパラメータ i を1だけインクリメントしてステップS1に戻り上記の処理を繰り返す。ステップS6でYESのときは、終了したと判断して当該MCE学習処理を終了する。ここで、学習データメモリ31内の学習データの一例を次の表に示す。

【0037】

【表4】

文章データ

/ 中間言語データ

八方尾根スキー場の標高差を教えて下さい。

/
R SHOWVALUE O_標高差 D_スキー場名 C_= V_八方尾
根

標高差が1000メートル以上のスキー場を教えて下さい。

/
R SHOWLIST O_スキー場名 D_標高差 C_>= V_100
0

【0038】図7は、図1の中間言語生成部21によって実行される中間言語生成処理を示すフローチャートである。図7において、まず、ステップS11でバッファメモリ24から音声認識結果を1文読み込む。次いで、ステップS12で音声認識結果の単語列W(L単語)に対して、次式を用いて中間言語の各要素 s_i ($1 \leq i \leq N$; Nは中間言語の要素数である。)毎に隠れマルコフモデル(HMM)による文生成確率 $P(W|s_i)$ を計算する。

【数8】
$$P(W|s_i) = \arg \max_L \prod_{x,y} a_{xy}(s_i) \cdot b_{y,w_i}(s_i)$$

ただし、 w_i は単語列Wの1番目の単語 a_{xy} (s_i)は中間言語の要素 s_i のモデルにおける、状態 x から状態 y への遷移確率 $a_{xy}(s_i)$ は中間言語の要素 s_i のモデルにおける、状態 y から単語 w_i への出力確率を表す。

【0039】次いで、ステップS13で中間言語の各要素の文生成確率 $P(W|s_i)$ に対して、各グループ内で尤度最大のものを選択する。すなわち、中間言語の各グループ(R_*, O_*, D_*, C_*, V_*)

において、そのグループに属する中間言語要素の内、ステップS12で求めた文生成確率 $P(W|s_i)$ の最も高いものを選択する。次いで、ステップS14で選択された中間言語の要素を所定のフォーマットにしたがって並べた後出力する。すなわち、ステップS13で入力された中間言語の各要素を中間言語文のフォーマット、すなわち、R_*, O_*, D_*, C_*, V_*の順番に従って並べ、中間言語を生成してコマンド実行部22に出力する。さらに、ステップS15で処理すべき音声認識結果がまだあるか否かが判断され、YESのときはステップS11に戻り上記の処理を繰り返す。一方、NOのときは当該中間言語生成処理を終了する。

【0040】図8は、図1のコマンド実行部22によって実行されるコマンド実行処理を示すフローチャートである。図8において、まず、ステップS21で中間言語生成部21から中間言語を1文入力する。次いで、ステップS22で、中間言語をデータベース言語(SQL言語)に変換する。すなわち、この変換は、次の表に示すように機械的に行われる。

【0041】

【表5】

中間言語 :

"R_SHOWVALUE O_標高差 D_キー場名 C_=
V_八方尾根"

データベース言語 :

"SELECT 標高差 FROM スキー場データ
WHERE スキー場名 = 八方尾根"

【0042】ここで、データベース言語は、「SELECT (1) FROM スキー場データ WHERE (2) (3) (4)」の形式をしており、(1)から(4)までの項目を、それぞれ、中間言語のO_、D_、C_、V_等の頭文字を削除した物を並べることにより用意に変換が可能である。

【0043】次いで、ステップS23で変換されたデータベース言語に基づいて、所定の条件に適合するデータ

をデータベースメモリ23から検索する。さらに、ステップS24では、データベースより得られたデータを中間言語のコマンド名に応じて加工して出力する。すなわち、ステップS23で得られた値を中間言語のコマンド名に応じて表示する。具体的には、次の表に示す表示処理を行う。

【0044】

【表6】

コマンド名

→

表示内容

R_SHOWVALUE	→	データベースから得られた値を表示
R_SHOWLIST	→	データベースからスキー場名の一覧を表示
R_SHOWIMAGE	→	データベースから得られたファイル名の画像を表示

【0045】ステップS24における表示処理においては、表示内容のデータがコマンド実行部22からパーソナルコンピュータ40を介してCRTディスプレイ41に出力されて表示され、また、表示内容のデータがプリンタ42に出力されて印字される。さらに、ステップS25では、処理すべき中間言語があるか否かが判断され、YESのときはステップS21に戻り上記の処理を繰り返す一方、NOのときは当該コマンド実行処理を終了する。

【0046】図1において、MCE学習処理部30、中間言語生成部21、及びコマンド実行部22は例えばデジタル計算機で構成され、学習データメモリ31、隠れマルコフモデル(HMM)メモリ32、バッファメモリ24及びデータベースメモリ23は例えばハードディスクメモリなどの記憶装置で構成される。

【0047】図2に本実施形態で用いる連続音声認識装置10のブロック図を示す。本実施形態の連続音声認識装置10は、公知のワンーパス・ビタビ復号化法を用いて、入力される発聲音声文の音声信号の特徴パラメータに基づいて上記発聲音声文の単語仮説を検出し尤度を計算して出力する単語照合部4を備えた連続音声認識装置において、単語照合部4からバッファメモリ5を介して出力される、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、統計的言語モデル13を参照して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する1つの単語仮説で代表させるように

単語仮説の絞り込みを行う単語仮説絞込部6を備える。

【0048】ここで用いる統計的言語モデル13は、学習用テキストデータに基づいて言語モデル生成部(図示せず。)により生成されたものであって、統計的言語モデル13は、例えば特開平9-134192号公報において開示されたように、品詞クラス間のバイグラム(N=2)を基本としたものであるが、単独で信頼できる単語は品詞クラスより分離させ、単独のクラスとして取り扱い、さらに、予測精度を向上させるため、頻出単語列に関してはそれらの単語を結合して一つのクラスとして取り扱い、長い単語鎖鎖の表現を可能にさせ、こうして、生成されたモデルは、品詞バイグラムと可変長単語N-グラムとの特徴を併せ持つ統計的言語モデルとなり、遷移確率の精度と信頼性とのバランスをとられたものである。

【0049】図2において、単語照合部4に接続され、例えばハードディスクメモリに格納される音素HMM11は、各状態を含んで表され、各状態はそれぞれ以下の情報を有する。

- (a) 状態番号
- (b) 受理可能なコンテキストクラス
- (c) 先行状態、及び後続状態のリスト
- (d) 出力確率密度分布のパラメータ
- (e) 自己遷移確率及び後続状態への遷移確率

なお、本実施形態において用いる音素HMM11は、各分布がどの話者に由来するかを特定する必要があるため、所定の話者混合HMMを変換して生成する。ここ

で、出力確率密度関数は34次元の対角共分散行列をもつ混合ガウス分布である。また、単語照合部4に接続され、例えばハードディスクに格納される単語辞書12は、音素HMM11の各単語毎にシンボルで表した読みを示すシンボル列を格納する。

【0050】図2において、話者の発声音声はマイクロホン1に入力されて音声信号に変換された後、特徴抽出部2に入力される。特徴抽出部2は、入力された音声信号をA/D変換した後、例えばLPC分析を実行し、対数パワー、16次ケプストラム係数、△対数パワー及び、16次△ケプストラム係数を含む34次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列はバッファメモリ3を介して単語照合部4に入力される。単語照合部4は、ワンーパス・ビタビ復号化法を用いて、バッファメモリ3を介して入力される特徴パラメータのデータに基づいて、音素HMM11と単語辞書12とを用いて単語仮説を検出し尤度を計算して出力する。ここで、単語照合部4は、各時刻の各HMMの状態毎に、単語内の尤度と発声開始からの尤度を計算する。尤度は、単語の識別番号、単語の開始時刻、先行単語の違い毎に個別にもつ。また、計算処理量の削減のために、音素HMM11及び単語辞書12とに基づいて計算される総尤度のうちの低い尤度のグリッド仮説を削減する。単語照合部4は、その結果の単語仮説と尤度の情報を発声開始時刻からの時間情報（具体的には、例えばフレーム番号）とともにバッファメモリ5を介して単語仮説絞込部6に出力する。

【0051】単語仮説絞込部6は、単語照合部4からバッファメモリ5を介して出力される単語仮説に基づいて、統計的言語モデル13を参照して、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する1つの単語仮説で代表させるように単語仮説の絞り込みを行った後、絞り込み後のすべての単語仮説の単語列のうち、最大の総尤度を有する仮説の単語列を認識結果として、バッファメモリ24を介して音声言語理解装置20の中間言語生成部21に出力し、上述の音声言語理解装置20の処理が実行される。本実施形態においては、好ましくは、処理すべき当該単語の先頭音素環境とは、当該単語より先行する単語仮説の最終音素と、当該単語の単語仮説の最初の2つの音素とを含む3つの音素並びをいう。

【0052】例えば、図3に示すように、(i-1)番目の単語 W_{i-1} の次に、音素列 a_1, a_2, \dots, a_n からなるi番目の単語 W_i がくるときに、単語 W_{i-1} の単語仮説として6つの仮説 $W_a, W_b, W_c, W_d, W_e, W_f$ が存在している。ここで、前者3つの単語仮説 W_a, W_b, W_c の最終音素は/x/であるとし、後者3つの単語仮説 W_d, W_e, W_f の最終音素は/y/であるとす

る。終了時刻 t_e と先頭音素環境が等しい仮説（図3では先頭音素環境が“x/a₁/a₂”である上から3つの単語仮説）のうち総尤度が最も高い仮説（例えば、図3において1番上の仮説）以外を削除する。なお、上から4番めの仮説は先頭音素環境が違うため、すなわち、先行する単語仮説の最終音素がxではなくyであるので、上から4番めの仮説を削除しない。すなわち、先行する単語仮説の最終音素毎に1つの仮説を残す。図2の例では、最終音素/x/に対して1つの仮説を残し、最終音素/y/に対して1つの仮説を残す。

【0053】以上の実施形態においては、当該単語の先頭音素環境とは、当該単語より先行する単語仮説の最終音素と、当該単語の単語仮説の最初の2つの音素とを含む3つの音素並びとして定義されているが、本発明はこれに限らず、先行する単語仮説の最終音素と、最終音素と連続する先行する単語仮説の少なくとも1つの音素とを含む先行単語仮説の音素列と、当該単語の単語仮説の最初の音素を含む音素列とを含む音素並びとしてもよい。

【0054】図2において、特徴抽出部2と、単語照合部4と、単語仮説絞込部6と、言語モデル生成部20とは、例えば、デジタル電子計算機で構成され、バッファメモリ3、5は例えばハードディスクメモリなどの記憶装置で構成され、音素HMM11と単語辞書12と統計的言語モデル13とは、例えばハードディスクメモリなどの記憶装置に記憶される。

【0055】以上実施形態においては、単語照合部4と単語仮説絞込部6とを用いて音声認識を行っているが、本発明はこれに限らず、例えば、音素HMM11を参照する音素照合部と、例えばOne Pass DPアルゴリズムを用いて統計的言語モデル13を参照して単語の音声認識を行う音声認識部とで構成してもよい。

【0056】

【実施例】本発明者は、音声言語理解装置20における言語理解率を評価するために、まず、正解文からの言語理解率を評価した。実験に用いたデータは、本特許出願人が所有するスキー場案内システムのために収集している会話で、現在、443文、7,569単語あり、語彙は281語である。全ての文章に、それに対応する中間言語を人手で作成している。言語理解のためのモデルは、最尤推定による隠れマルコフモデル（ML-HMM）、及び、最尤推定後にMCE学習を行ったモデル（MCE-HMM）の2種類用意した。ただし、隠れマルコフ状態数は、いずれのモデルも2とした。

【0057】評価は言語理解率で行った。ただし、言語理解率は、入力文章から中間言語へ正確に変換できた割合であり、中間言語の全ての要素が正しく変換できた場合のみ正解とする。最尤推定による隠れマルコフモデル（HMM）を用いた場合、言語理解率は96.0%となり高い率を得た。さらにMCE学習を行うことにより

言語理解率は99.6と極めて高い率を得ることができた。

【0058】以上説明したように、本実施形態によれば、隠れマルコフモデルを用いた統計的手法により、自然言語から中間言語への変換を行い、言語理解を行う音声言語理解システムを構築して、最尤推定による隠れマルコフモデルにおける言語理解率が96.0%であり、さらにMCE学習を行うことにより、99.6%と非常に高い確率で言語理解率が得られることを確認した。音声言語理解装置20は、統計的手法を用いて処理するため、文法ルールの作成やキーワードの選択等の作業を必要とせず、また、モデルの学習には数千語程度のデータで良好な結果を得るため、短時間でのシステム構築が可能であるという利点がある。すなわち、従来技術に比較して頑健にかつ正確に音声理解を行うことができ、適切に応答する処理を実行することができる音声言語理解装置20及び、音声言語理解システムを提供することができる。

【0059】

【発明の効果】以上詳述したように、本発明に係る請求項1記載の音声認識装置によれば、発聲音声から音声認識装置によって音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解装置であって、検索条件を含む文章データと、それに対応しかつ少なくとも応答する処理の内容及びデータの項目名を含む所定の中間言語データとの対である学習データに基づいて、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習して得られた隠れマルコフモデルを記憶する第1の記憶装置と、複数の項目名に対応したデータを含むデータベースを記憶する第2の記憶装置と、上記第1の記憶装置に記憶された隠れマルコフモデルを用いて、上記音声認識されかつ検索条件を含む音声認識結果の音声文を上記中間言語に変換して生成する生成手段と、上記生成手段によって生成された中間言語を、上記データベースに対応した所定のデータベース言語に変換した後、上記データベース言語に含まれる検索条件に基づいて、上記第2の記憶装置に記憶されたデータベースを検索して、上記検索条件を満たすデータを獲得し、そのデータについて上記中間言語に対応した応答する処理を実行する実行手段とを備える。従って、従来技術に比較して頑健にかつ正確に音声理解を行うことができ、適切に応答する処理を実行することができる音声言語理解システムを提供することができる。

【0060】また、本発明に係る請求項2記載の音声言語理解システムによれば、発聲音声を音声認識して、音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解シ

ステムであって、発聲音声を音声認識して、音声認識された音声認識結果の音声文を出力する音声認識装置と、検索条件を含む文章データと、それに対応しかつ少なくとも応答する処理の内容及びデータの項目名を含む所定の中間言語データとの対である学習データに基づいて、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習して得られた隠れマルコフモデルを記憶する第1の記憶装置と、複数の項目名に対応したデータを含むデータベースを記憶する第2の記憶装置と、上記第1の記憶装置に記憶された隠れマルコフモデルを用いて、上記音声認識装置によって音声認識されかつ検索条件を含む音声認識結果の音声文を上記中間言語に変換して生成する生成手段と、上記生成手段によって生成された中間言語を、上記データベースに対応した所定のデータベース言語に変換した後、上記データベース言語に含まれる検索条件に基づいて、上記第2の記憶装置に記憶されたデータベースを検索して、上記検索条件を満たすデータを獲得し、そのデータについて上記中間言語に対応した応答する処理を実行する実行手段とを備える。従って、従来技術に比較して頑健にかつ正確に音声理解を行うことができ、適切に応答する処理を実行することができる音声言語理解システムを提供することができる。

【図面の簡単な説明】

【図1】 本発明に係る一実施形態である音声言語理解装置20を備えた音声言語理解システムの構成を示すブロック図である。

【図2】 図1の音声認識装置10の構成を示すブロック図である。

【図3】 図2の音声認識装置における単語仮説絞込部6の処理を示すタイミングチャートである。

【図4】 図1の隠れマルコフモデル(HMM)メモリ32に格納されたHMMを示す状態遷移図である。

【図5】 図1の中間言語生成部21の処理を示す説明図である。

【図6】 図1のMCE学習処理部30によって実行されるMCE学習処理を示すフローチャートである。

【図7】 図1の中間言語生成部21によって実行される中間言語生成処理を示すフローチャートである。

【図8】 図1のコマンド実行部22によって実行されるコマンド実行処理を示すフローチャートである。

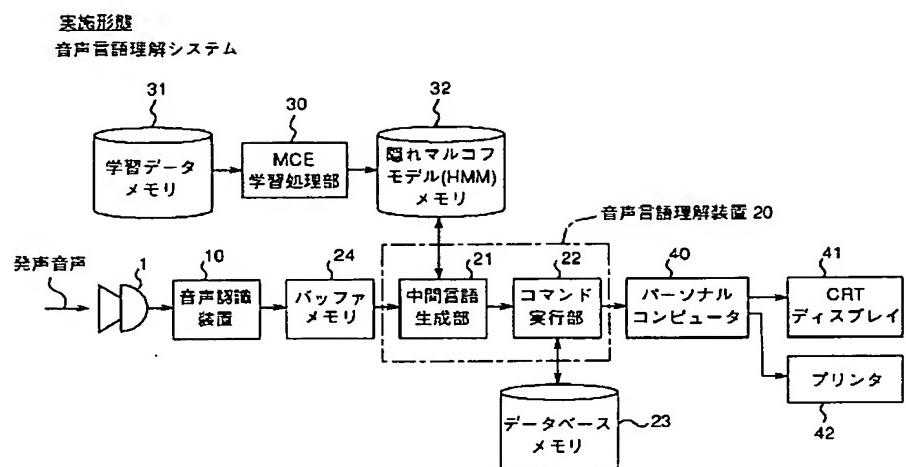
【符号の説明】

- 1…マイクロホン、
- 2…特徴抽出部、
- 3, 5…バッファメモリ、
- 4…単語照合部、
- 6…単語仮説絞込部、
- 1 1…音素HMM、
- 1 2…単語辞書、
- 1 3…統計的言語モデル、

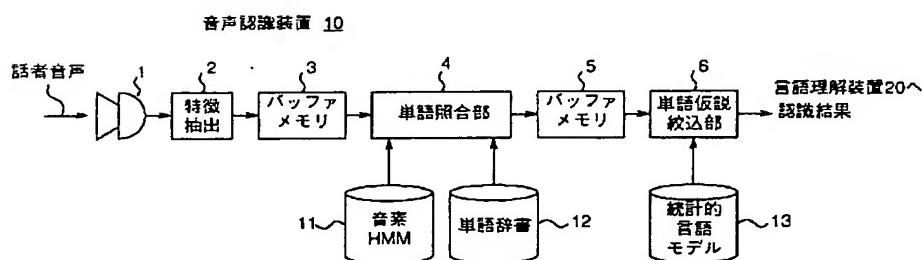
20…音声言語理解装置、
21…中間言語生成部、
23…データベースメモリ、
24…バッファメモリ、
30…MCE学習処理部、
31…学習データメモリ、
32…隠れマルコフモデル(HMM)メモリ。

31…学習データメモリ、
32…隠れマルコフモデル(HMM)メモリ、
40…パーソナルコンピュータ、
41…CRTディスプレイ、
42…プリンタ。

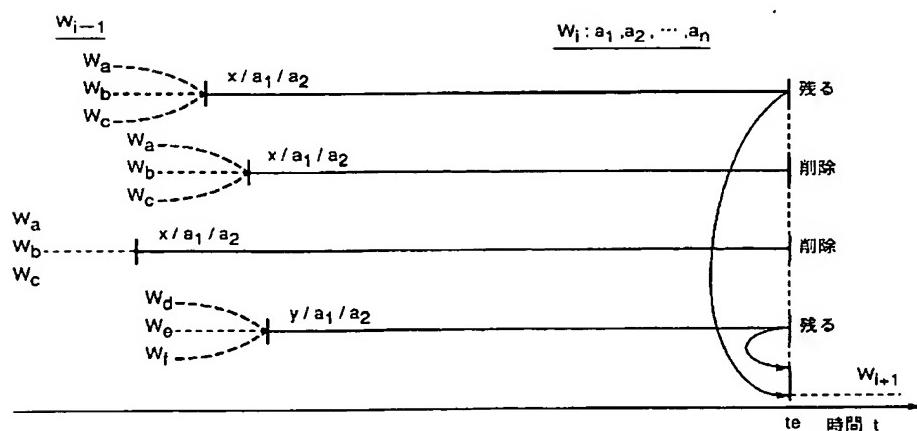
【図1】



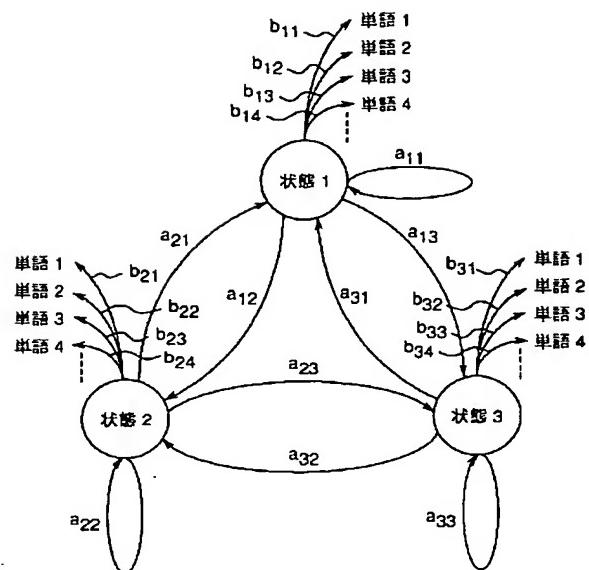
【図2】



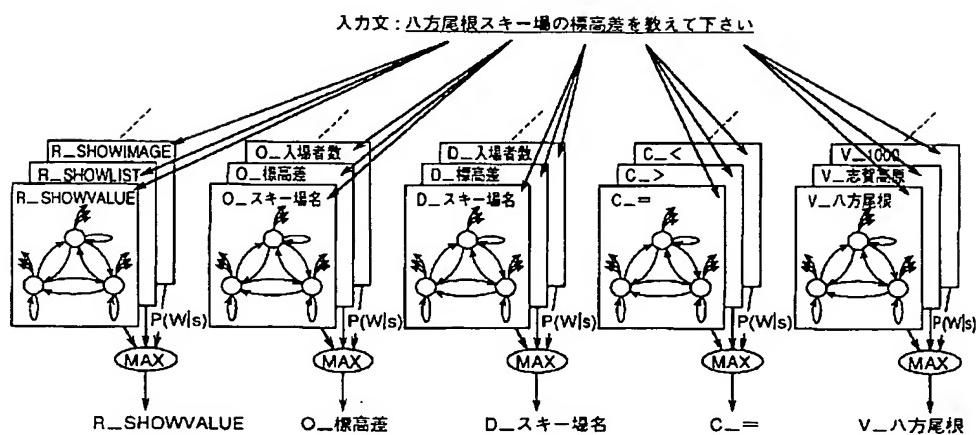
【図3】



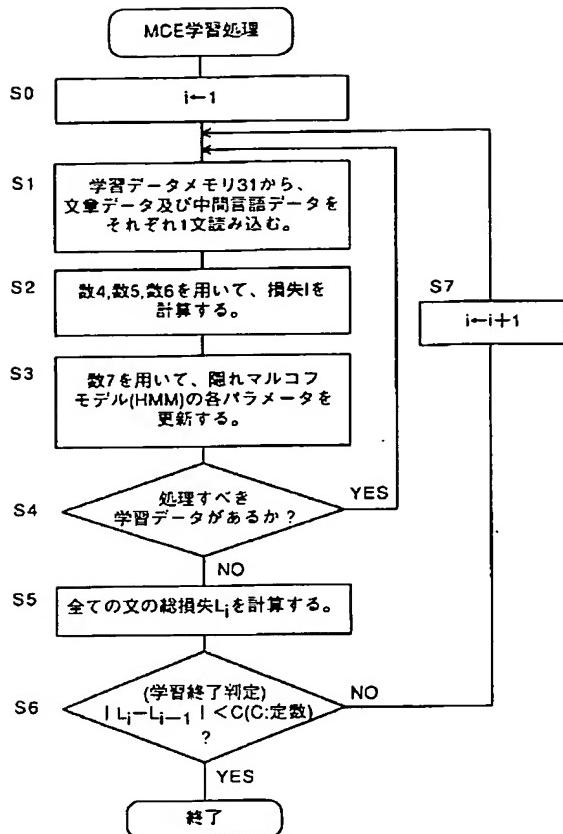
【図4】



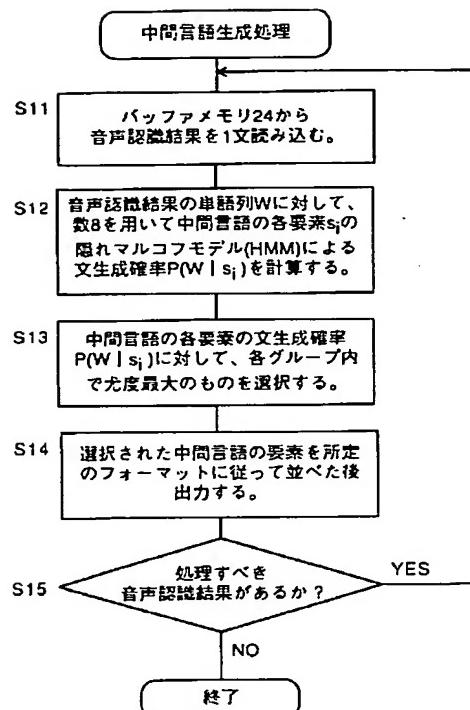
【図5】



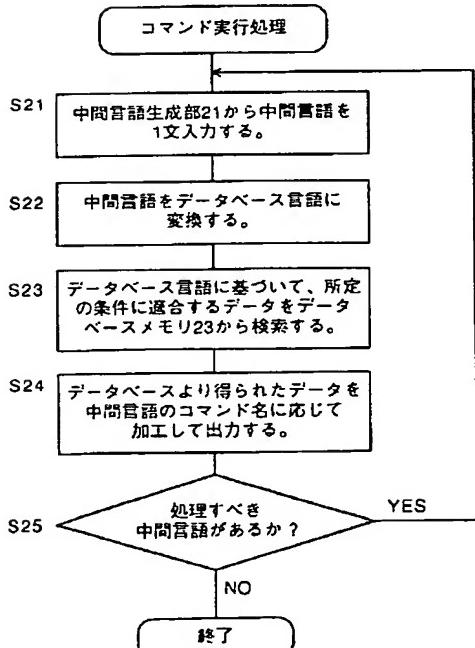
【図6】



【図7】



【図8】



フロントページの続き

(51) Int.C1.⁶
G 0 6 F 17/30
// G 0 6 F 3/16

識別記号
3 2 0

F I
G 0 6 F 15/38
15/403

P
3 1 0 Z

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-143493
 (43)Date of publication of application : 28.05.1999

(51)Int.CI. G10L 3/00
 G10L 3/00
 G10L 3/00
 G06F 15/18
 G06F 17/28
 G06F 17/30
 // G06F 3/16

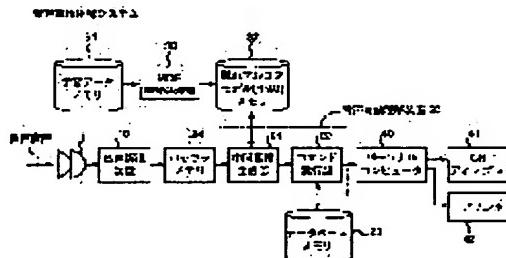
(21)Application number : 09-303075 (71)Applicant : ATR ONSEI HONYAKU TSUSHIN KENKYUSHO:KK
 (22)Date of filing : 05.11.1997 (72)Inventor : MASATAKI HIROKAZU

(54) DEVICE AND SYSTEM FOR UNDERSTANDING VOICE WORD

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a device and a system for understanding the voice word capable of firmly and correctly understanding the voice, and executing the appropriately responding processing compared with the conventional technology.

SOLUTION: An MCE learning part 30 learns a hidden Markov model to convert sentence data into intermediate words corresponding thereto so that the discrimination error is minimized based on the learned data. An intermediate word generation part 21 converts the voice sentence of the result of the speech recognition which is speech recognized and includes the retrieval condition using a hidden Markov model to convert the sentence data into the intermediate words corresponding thereto. After a command execution part 22 converts the generated intermediate words into the prescribed database words corresponding to the database, it retrieves the database based on the retrieval condition included in the database words, acquires the data to satisfy the retrieval condition, and executes the processing responding to the intermediate words on the data.



LEGAL STATUS

[Date of request for examination] 05.11.1997

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3088364
[Date of registration] 14.07.2000
[Number of appeal against examiner's decision
of rejection]
[Date of requesting appeal against examiner's
decision of rejection]
[Date of extinction of right] 14.07.2003

Copyright (C); 1998,2003 Japan Patent Office

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

BLACK BORDERS

IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

FADED TEXT OR DRAWING

BLURRED OR ILLEGIBLE TEXT OR DRAWING

SKEWED/SLANTED IMAGES

COLOR OR BLACK AND WHITE PHOTOGRAPHS

GRAY SCALE DOCUMENTS

LINES OR MARKS ON ORIGINAL DOCUMENT

REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.